DOCUMENT RESUME

ED 200 614 TM 810 159

AUTHOR Anderson, Richard Ivan
TITLE A PLATO Probabilistic Testing System.
PUB DATE 80
NOTE 16p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Data Collection; Feedback; *Probability; *Response Style (Tests); *Scoring Formulas; Test Items
IDENTIFIERS *PLATO; *TUTOR Programing Language

ABSTRACT

Features of a probabilistic testing system that has been implemented on the "cerl" PLATO computer system are described. The key feature of the system is the manner in which an examinee responds to each test item; the examinee distributes probabilities among the alternatives of each item by positioning a small square on or within an equilateral triangle. Each of the three vertices corresponds to an item alternative. Responses are scored using an "admissible" probabilistic scoring function which enables examinees to maximize their scores in the long run if they honestly assign probabilities to alternatives. The difference between the score obtained by an examinee and the maximum possible score can be divided into two parts: (1) gain possible via a gain in knowledge about the topics covered by the test, and (2) gain possible via more appropriate assignments of probabilities to alternatives. The testing system splits the difference in this manner and reports the information to the examinee. Specific topics covered in the paper include: (1) triangle response aid; (2) scoring function; (3) performance feedback; (4) before, during, and after test administration; (5) restrictions placed on tests; (6) system TUTOR code; (7) system use; and (8) data collection. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Anderson RI

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RECEIVED
ERIC/TM

ED200614

TM 810 159

A PLATO Probabilistic Testing System

Richard Ivan Anderson

1980

The purpose of this paper is to describe features of a probabilistic testing system[1] that has been implemented on the "cerl" PLATO computer system. The antecedent of this probabilistic testing system has been described in varying detail by others (Landa, 1976; Shuford, 1974; Sibley, 1973; Rippey, Note 1; Rippey & Smith, 1979). Contrasts between the present system and its antecedent will be drawn where appropriate.

Key Features

The key feature of the system is the manner in which an examinee responds to each test item. An examinee distributes probabilities among the alternatives of each item by positioning a small square on or within an equilateral triangle. Since each vertex corresponds to an item alternative, the number of alternatives per item is limited to three.

Responses are scored using an "admissible" probabilistic scoring function (Shuford, Albert, & Massengill, 1966). An admissible probabilistic scoring function supposedly enables examinees to maximize their scores in the long run if and only if they honestly assign probabilities to alternatives. To reflect honest responding, alternatives assigned a particular probability must be correct a proportion of times equal to that particular probability. However, this relationship can occur when an examinee dishonestly assigns equal probability to all alternatives. Thus, this relationship, while not synonymous with honest responding, is supposedly the appropriate product of honest responding.

Under the assumption that alternatives assigned a particular probability should be correct a proportion of times equal to that particular probability, the difference between the score obtained by an examinee and the maximum possible score can be divided into two parts: 1) gain possible via a gain in knowledge about the topic(s) covered by the test; 2) gain possible via more appropriate assignments of probabilities to alternatives. The testing system splits the difference in this manner and reports the information to the examinee.

Triangle Response Aid

Displayed at each vertex are the corresponding alternative, the probability assigned to the corresponding alternative (initialized to .33), and the score the examinee will receive if the corresponding alternative is correct (initialized to 0). The antecedent system did not display the three probability values. However, some evidence suggests that the unusual values produced by admissible scoring functions can inappropriately influence examinee responses (Jensen & Peterson, 1973). Under the speculation that examinee responses may be inappropriately influenced more readily if the scores were displayed without the probability values, probabilities were added to the vertices.

An examinee positions the small square on or within the triangle by pressing arrow keys available on a PLATO keyset. Each press of an arrow key is followed by either a repositioning of the square and a value change in the probabilities and scores displayed at the vertices or a message specifying that the pressed arrow key can't be used at that point. The

latter will occur either when the pressed arrow key directs the square outside of the triangle's boundaries or for a subset of arrow keys when the square is positioned at the triangle's center.

Probabilities that can be assigned are restricted to the following set: "0", .1, .2, .3, .33, .4, .5, .6, .7, .8, .9, 1. The antecedent system allowed the assignment of any two-decimal-place probability value. However, the mechanism via which examinees responded in the antecedent system sometimes impeded positioning the square precisely. The restrictions placed on the assignable probabilities enabled making the task of positioning the square more straightforward while not limiting assignable probabilities excessively.

The zero probability value above is surrounded by quotation marks since the smallest probability value that can be assigned is actually .01. This is true since, as is described later, the admissible probabilistic scoring function can not accomodate a probability value of zero. Because of the need to make the smallest assignable probability value non-zero, the sum of the probabilities assigned to the three alternatives will not always be precisely one (e.g., .6+.4+.01; .01+.01+1). Since these inequalities are not revealed to examinees, their existence is not considered to be a problem.

Scoring Function

The admissible probabilistic scoring function used is $50 \times \log(3p_c)$, where p_c = probability assigned by the examinee to the correct alternative. Scores are rounded to the nearest integer in all cases. Although rounding

introduces some degree of error into the scores, it seemingly eliminates examinee confusion that fractional values might generate. in examinees. The antecedent system rounded scores to the nearest integer only for the displays at triangle vertices; an examinee's total score reflected all fractional values.

The logarithmic scoring function reveals why the smallest assignable proability value can not be zero. The common log of zero is negative infinity.

Performance Feedback

To begin the process of dividing the difference obtained between the score received by an examinee and the maximum possible score into the gain possible via a gain in knowledge and the gain possible via more appropriate assignments of probabilities, a least-squares regression line constrained to the point (1/3, 1/3) is calculated that describes the relationship between the assigned probabilities and the proportions of times the alternatives to which they were assigned were correct. Then, the test is rescored by substituting predicted proportions of times correct for probabilities actually assigned. These substitutions correct assigned probabilities in accordance with the presumed most appropriate relationship between assigned probabilities and the proportions of times the alternatives to which they are assigned are correct. Since a perfect relationship theoretically maximizes an examinee's total score, the rescore should more often than not be higher than the original score. The difference between the rescore and the original score represents the gain possible via more appropriate

assignments of probabilities to alternatives. The difference between the rescore and the maximum possible score represents the gain possible via a gain in knowledge about the topic(s) covered by the test.

Revelation of the divided difference values to an examinee is accompanied by a visual of the relationship between the amount of knowledge that the examinee's responses suggested that he or she thought he or she possessed and the amount of knowledge that the examinee's responses suggested that he or she actually possessed. Also presented is a verbal assessment of the same relationship.

Additional Student-System Interaction Features

Student-system interaction features additional to those described above are activated before test administration, during test administration, and after test administration. The primary additional features are described below.

Before Test Administration

Two short tutorials can be viewed at an examinee's option before he or she takes a test. One describes probabilistic testing in general; one describes details of responding to a probabilistic test on PLATO. Both are additions to the antecedent system.

Brief instructions on how to respond to a probabilistic test on PLATO are automatically presented prior to each test administration.

During Test Administration

An examinee is allowed to progress linearly, forward and backward, through a test and, as desires, adjust a response to any test item. The antecedent system allowed only a forward linear progression through the test and, consequently, allowed no option to adjust earlier responses. However, the antecedent system did give an examinee the option to learn which alternative was correct after responding to each item. This feature has been disabled.

An option to review instructions on how to respond to a probabilistic test on PLATO was unavailable in the antecedent system but is available now. After this option is selected, the brief instructions that are automatically presented prior to test administration are presented automatically again, and the examinee is given the option to examine the more detailed tutorial. Return to the test occurs at the examinee's request and is made to the point of departure from the test.

The scoring of the entire test can be requested at any time by the examinee. But, if any test item has not been presented to the examinee, the scoring request is denied, and the first of the items not yet presented is displayed.

After Test Administration

After the performance feedback described earlier is presented to the examinee, he or she is given an option not available in the antecedent system to review the test. During the review, correct alternatives are identified. An examinee can progress linearly, forward and backward, through the test as long as desired.

Another feature unavailable in the antecedent system gives the examinee the option to examine an explanation of the performance feedback. Selection of this option is followed by the presentation of a brief tutorial.

Two additional options are available: 1) return to the test index; 2) depart from the testing system.

Additional Information

Additional, relevant information is described below.

Mechanism for Presenting Test Items

The order of test item presentation and the positioning of alternatives at vertices is random.

Restictions Placed on Tests

The testing system can currently accomodate a maximum of 5 tests, each test of 50 or fewer items, at one time. The antecedent system unnecessarily restricted test size to 9 items.

Test items are limited in size due to a finite PLATO screen space. Each item stem can occupy no more than 5 lines, each line of 59 characters. Each of the three item alternatives can occupy no more than 3 lines, each line of 26 characters. As suggested by Rippey (Note 1), needs for larger item sizes than allowed by the system can be met by placing the items on paper. However, the random order of test item presentation and alternative positioning somewhat hinders the implementation of such a plan.

System TUTOR Code

The testing system's TUTOR code differs dramatically from the TUTOR code of the antecedent system. When received by the author,[2] the code of the antecedent system was in bad shape.

The system currently occupies a portion of a 3-part PLATO lesson space. Three-part PLATO lesson spaces occupy roughly 6700 60-bit words. The test system machinery including in-code documentation and the three explanatory tutorials, but not including the test items, occupies roughly 4100 of those 6700 words. Half of the 4100 words are occupied by the tutorials.

System Use

To -use- the system, obtain your own lesson space and specify in its directory that it -use-s lesson "anderson". Then, insert the TUTOR code appearing in Figure 1.

Insert Figure 1 about here

Replace the "?" in the -define- statement by the number of tests that will be made available at one time to your examinees. If you plan to make available only 1 test, test information must be added to -unit-s "title1", "total1", "ques1", "1corr", "1wrg1", and "1wrg2". If you plan to make available only 2 tests, test information must be added to -unit-s "title1", "total1", "ques1", "1corr", "1wrg1", "1wrg2", "title2", "total2", "ques2", "2corr", "2wrg1", and "2wrg2". If you plan to make available only 3 tests, test information ... Even if you make available less than 5 tests, all of the -unit-s for all 5 tests must appear in your lesson space.

To make a test available, follow the instructions below:

1) Append a test title to the -write- statement in the "title" -unit-. (Note the specified maximum length.)

2) Replace the "?" in the "total" -unit-'s -calc- statement with the number of test items. (Note the specified maximum number.)

3) Place item stems in the -writec- statement in the "ques" -unit-. (Note the specified maximum size.) Consecutive stems should be separated by a "$". (Stems could be read from a dataset rather than placed in the TUTOR code.)

4) Place item alternatives in the -writec- statements in the "corr" and "wrg" -unit-s. (Note the specified maximum size.) The correct alternatives should be placed in the "corr" -unit-; the wrong alternatives should be placed in the "wrg" -unit-s. Consecutive alternatives should be separated by a "$". Order of alternatives within -unit-s should correspond to the order of item stems within the "ques" -unit-. (Alternatives could be read from a dataset rather than placed in the TUTOR code.)

Because various aspects of the system may change, contact the author before -use-ing. Use of the system has been restricted to those who have the appropriate codeword in order to ensure an author contact is made.

Data Collection

To save examinee scores, replace the "0" in the -define- statement appearing in Figure 1 with "1", and add TUTOR code to -unit- "savedata" that will save an examinee's scores.

An examinee's scores appear in two variables: 1) "tot" = total score obtained by examinee; 2) "retot" = total score obtained when assigned probabilities were adjusted to more appropriate values.

Relevant arithmetic expressions: 1) "retot-tot" = gain possible via more appropriate assignments of probabilities to alternatives; 2) "24xprobtot-retot" = gain possible via a gain in knowledge about the topic(s) covered by the test.

Reference Note

1. Rippey, R. M. Instructions for utilizing confidence testing on the PLATO system. Unpublished manuscript, undated.

References

Jensen, F. A. & Peterson, C. R. Psychological effects of proper scoring rules. Organizational Behavior and Human Performance, 1973, 9, 307-317.

Landa, S. CAAPM: Computer-aided admissible probability measurement on PLATO IV (Report No. 1721-ARPA). Santa Monica, CA: Rand Corporation, March, 1976. (ERIC Document Reproduction Service No. ED129850)

Shuford, E. H. The student as an assessor of uncertainty: Some statistical measures useful for feedback to the student. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1974. (ERIC Document Reproduction Service No. ED093995)

Shuford, E. H., Albert, A., & Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 125-145.

Sibley, W. L. An experimental implementation of computer assisted admissible probability testing. Paper presented at the Military Association Conference, San Antonio, 1973. (ERIC Document Reproduction Service No. ED104901)

Rippey, R. M. & Smith, S. Improving the reliability and validity of confidence-scored tests by adjusting for realism. Evaluation and the Health Professions, 1979, 1, 100-109.

Footnotes

[1]Various aspects of the system are subject to change. To learn of any changes that may have been made which falsify any part of this paper, contact the author.

[2]The antecedent system was made available to the author via the courtesy of Robert M. Rippey.

Figure Caption

Figure 1. TUTOR code that must be inserted into a lesson space in order to -use- the probabilistic testing system.

```
define  qavail=?
        datasav=0
use     ieu
use     main
use     scoring
use     help
use     arrows
unit    savedata
*
unit    titlel
* 44 character maximum
write
*
unit    totall
* 50 item maximum
calc    probtot←?
*
unit    quesl
* 5 lines x 59 character maximum
writec  randnol-2⇕
*
unit    icorr
* 3 lines x 26 character maximum
writec  randnol-2⇕
*
unit    lwrgl
* 3 lines x 26 character maximum
writec  randnol-2⇕
*
unit    lwrg2
* 3 lines x 26 character maximum
writec  randnol-2⇕
*
```

Repeat 4 times, first replacing the "1" in each -unit- name (left-most only in "lwrgl") by a "2", then by a "3", then by a "4", and lastly by a "5".